

1997

Answering the Bayesioskeptical Challenge

Richard D. Friedman

University of Michigan Law School, rdfrdman@umich.edu

Available at: <https://repository.law.umich.edu/articles/730>

Follow this and additional works at: <https://repository.law.umich.edu/articles>



Part of the [Evidence Commons](#), and the [Jurisprudence Commons](#)

Recommended Citation

Friedman, Richard D. "Answering the Bayesioskeptical Challenge." *Int'l J. Evidence & Proof* 1 (1997): 276-91.

This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

Answering the Bayesioskeptical challenge

By Richard D. Friedman*

Professor of Law, University of Michigan Law School

In recent years, some scholars of evidence, myself among them, have made active use of subjective probability theory – what is sometimes referred to as Bayesianism – in thinking about issues and problems related to the law of evidence. But, at the same time, this use has been challenged to various degrees and in various ways by scholars to whom I shall apply the collective, if somewhat misleading, label of Bayesioskeptics.¹ I present this brief paper to defend this use of probability theory, and to discuss what I believe is its proper role in discourse about evidentiary issues.

The burdens of proof

A basic model

Mathematicians have long developed what might be termed conventional probability theory, in which probabilities obey certain basic principles, such as that the probability of proposition $A \& B$, given a body of information O , equals the probability of A , given O , times the probability of B , given A and O . Most people, I believe, understand the basic idea of probability with respect to propositions reflecting events that might recur in essentially identical form; thus, 'The probability that any given flip of a fair coin will land heads up is 1 in 2'. But what bearing does this have on litigation?

Most events or propositions of interest in litigation are not of the nature of a coin flip. Matters at issue in litigation tend to be one-time events. If Victor was murdered, it makes no sense to say, 'The probability that in any given murder of Victor the murderer would be Dennis is 1 in 2'. But assigning probabilities to propositions representing unique events does make sense if we treat those probabilities as nothing more than assessments of subjective levels of confidence in the truth of the

* Many thanks for comments and other assistance to Colin Aitken, Ron Allen, David Balding, Craig Callen, Peter Donnelly, David Kaye and Dennis Lindley.

1 There are variations among Bayesioskeptics, as among Bayesians. Some Bayesioskeptics are rather uncompromising; for example, Alex Stein, in messages sent to an internet discussion list while this article was in draft, has contended that subjective probability theory is 'vacuous' and lacks even heuristic value. Other Bayesioskeptics acknowledge some limited usefulness for Bayesian methods but spend considerable energy criticising those methods and little or none using them. See, for example, Callen (1982: 44).

proposition.² Thus, a person might say, 'I assess the probability that Dennis was the murderer of Victor as 1 in 2. If I had to bet, with the bet to be decided by someone who happens to be omniscient, I would be just as willing to bet on the proposition that Dennis was the murderer as I would to bet – with the same stakes – on the next flip of a fair coin being heads up'. Application of the conventional probability theory to subjective probability assignments is the essence of Bayesianism.³

Of course, in litigated matters there is no one omniscient around; if there were, we would not have to litigate. But the outcome of the adjudication will have consequences, for good or ill, even if we do not know when a case is decided which type of consequence the decision will have. So we are forced to decide under uncertainty, and the uncertainty is of a subjective nature; that is, different fact-finders may rationally make different probability assessments if for no other reason than that they approach the problem with different bases of information from the outside world. Thus, whatever the value of Bayesian methods as opposed to classical statistical methods in scientific inquiry, in litigation I believe that a subjectivist approach to probability is the only one that can offer any hope of assisting in the analysis of juridical proof. In the law, we simply do not have the luxury of casting aside subjective probability assessments and declining to reject a given hypothesis until we receive evidence that would be extremely unusual under that hypothesis.

Suppose for simplicity that we are limited to two options – finding for the plaintiff or finding for the defence. Then, at least as a first approximation, it seems that a wise guide to decision is to choose the option that has the greater expected utility (Lempert, 1977: 1021, 1032–41). Expected utility will depend on the relative probabilities that the facts support the plaintiff and the defence and on the utilities attached to each possible decision given each possible actual state of affairs. Thus:

$$EU(p) = P(\Pi) \times U(p, \Pi) + P(\Delta) \times U(p, \Delta), \quad (1)$$

and

$$EU(d) = P(\Pi) \times U(d, \Pi) + P(\Delta) \times U(d, \Delta), \quad (2)$$

where $EU(p)$ and $EU(d)$ represent the expected utilities of judgments for the plaintiff and the defendant, respectively, $P(\Pi)$ represents the probability that the facts are such that the plaintiff is entitled to judgment, and $P(\Delta)$ represents the comparable probability with respect to the defendant; the 'i' in $U(i, \sigma)$ represents the party who receives the judgment, and the σ represents the party who is in fact entitled to judgment. Thus, for example, $U(p, \Delta)$ equals the social utility of a judgment for the

2 Probability theory does not purport to prescribe or confine the probability assessments that an observer initially makes, but given those initial assessments the theory does impose consistency conditions. Here is a simple illustration. Consistently with the theory an observer might assign any probability to the proposition that it will be cloudy in the morning or to the proposition that it will be rainy in the afternoon. But if she deems that it is improbable that it will be cloudy in the morning she could not deem it probable that it will be both cloudy in the morning and rainy in the afternoon.

3 Some theorists, for example, de Finetti (1974–5), contend that all probability assignments are of this subjective nature. Bayesianism is in a sense a misleading term: Bayes' theorem, an expression for updating probability assessments in light of further information, does not apply only to subjectively assigned probabilities, and there is much more to subjective probability theory than the theorem. But the theorem is crucial, and the label is handy.

plaintiff when the truth, if it were known, is such that the defendant should receive judgment. $U(p, \Pi)$ and $U(d, \Delta)$ must each have greater utility than $U(p, \Delta)$ and $U(d, \Pi)$; it is helpful to assume that the first pair has positive utility and the second pair has negative utility.

Now, at some degree of confidence in the proposition that the facts favour the plaintiff, $EU(p) = EU(d)$. This degree of confidence is the standard of persuasion. Bearing in mind that $P(\Pi) = 1 - P(\Delta)$, the standard of persuasion might be expressed in odds form as follows:

$$O(\Pi) = \frac{P(\Pi)}{1 - P(\Pi)} = \frac{U(d, \Delta) - U(p, \Delta)}{U(p, \Pi) - U(d, \Pi)} \quad (3)$$

Thus, judgment for the plaintiff will be optimal only if the fact-finder's degree of confidence in the plaintiff's case is at least as great as this level. As a reflection, perhaps, of inertia built into the system, the defendant wins if the fact-finder's level of confidence falls precisely at this level.

In a criminal case, the long-standing and solidly established view is that $U(p, \Delta)$, the negative utility of an incorrect judgment for the prosecution, far exceeds any of the other utilities in magnitude; Blackstone's injunction that it is better to let 10 guilty persons go free than to convict one innocent person may well be a substantial understatement. This utility assessment leads to the 'beyond a reasonable doubt' standard in criminal cases.

In civil cases, the usual conception is that a correct judgment for the plaintiff is about as good as a correct judgment for the defendant, and that an incorrect verdict for one is about as bad as an incorrect verdict for the other: that is, $U(p, \Pi) = U(d, \Delta)$, and $U(p, \Delta) = U(d, \Pi)$. This means that the standard of persuasion, expressed in odds, equals 1, or 0.5 expressed as a probability. This, of course, is the familiar 'more likely than not', or 'balance of probabilities', standard.

Now that seems straightforward enough. Determining a standard of persuasion is, it seems to me, an inevitable feature of litigation under uncertainty, and if we are to do it we need a concept of probability. In this context the charge is sometimes made of 'Bayesian imperialism'. To this charge, I think, Bayesians should plead half-guilty.⁴ I do not believe that alternative systems of probability – which would lead a fact-finder to results inconsistent with those she would reach by applying conventional probability – are useful in this context.⁵ The conventional theory is not contrived; it reflects rational treatment of uncertainty based on intuitively appealing and experientially successful premises.

4 One leading Bayesian has adopted the imperialistic tag as a badge of honour (Edwards, 1991: 1025, 1056).

5 Compare the comment of Alex Stein, in Allen *et al.* (1995: 285) ('"Imperialistic Bayesianism" refers to a theory that attempts to exclude a Baconian type of reasoning from the domain of rationality. One may reason about uncertainties by assessing the strength of the evidence which supports or weakens one's hypothesis, etc., as Jonathan Cohen explains.'). Assessments of probability under the conventional system are sensitive to the strength of evidence. I do agree with Stein that the strength of evidence may play a role, not fully captured in the probability assessment, in deciding whether a party should be entitled to judgment. But this is no reason to reject the applicability of the conventional theory. The Baconian system advocated by Cohen does not seem useful in the legal context. In Friedman (1996: 1820) I have stated several reasons why.

An only partially imperialistic view: probability as an incomplete determinant of the standard of persuasion

To assert the pre-eminence of conventional probability theory is not to claim that the simple model I have presented above is complete, and that the standard of persuasion can be defined across a wide range of cases in terms of a simple probabilistic assessment. Let me illustrate with an old chestnut, the blue bus hypothetical (e.g. Brook, 1985: 293). A woman has been run off the road by a bus that she cannot identify. The only information presented to help determine who should be liable is proof that the Blue Bus Company owns 80 per cent of the buses in town. Should the Blue Bus Company be liable?

First, let's clear away one misconception that sometimes enters into discussion of problems like this:⁶ The 80 per cent figure is not a probability, but only a datum. Probability assessments are made by observers on the basis of all the data received – which include information about what is *not* received. Sometimes the data received are so crisp and strong that they narrowly confine the range of probability assessments that most rational observers could make, but that is not this case. Failure of either side to produce more evidence might lead a rational fact-finder to assess a probability substantially higher or lower than the 80 per cent figure (Kaye, 1982).

I believe that part of the problem in this case may be that, despite the received understanding of the standard of persuasion, in fact we apply a standard much more stringent than a probability of 0.5 when the question is whether the defendant is the liable party or, alternatively, was not involved at all. Putting that aside, though, many people seem to have a strong visceral reaction against allowing a judgment for the plaintiff in the Blue Bus Case because the evidence is so skimpy. Even given the same probability of liability, most people have a much less satisfied feeling granting relief in a case in which the evidence is very incomplete, even if justifiably so, so that it is difficult to fill in a detailed picture of what happened. So we might say that the less complete the evidence the greater the probability must be to warrant relief for the plaintiff. In other words, instead of a simple rule of the form $P > K$, where P is the probability of liability and K the designated standard of persuasion, we might have a somewhat more complex rule of the form $P > K/Q$, where Q is some measure of the quality or completeness of the evidence. (I make this argument at somewhat greater length in Allen *et al.*, 1995: 309–10.)

In short, I do not believe that the standard of persuasion necessarily can be expressed solely in terms of probability. But probability, that is, the conventional theory of probability, clearly should play a large role in defining that standard.

The so-called problem of conjunction

The apparent problem

Some Bayesioskeptics have contended that the 'problem of conjunction' reflects a flaw inherent in the use of probability theory to define the standard of persuasion (Cohen, 1977: 58–67; Allen, 1986: 405–7; Allen, 1994: 605–12; Brilmayer and Kornhauser, 1978: 145–6).⁷ The supposed problem arises from the fact that a plaintiff must prove several factual propositions, or *elements*, to prevail on his claim. For example, in an ordinary torts case the plaintiff may have to prove: (1) that the defendant had a duty of care to the plaintiff, (2) that the defendant breached that

⁶ For example, Wright (1988: 1058); Allen *et al.* (1995: 291–2).

⁷ Cf. Nesson (1985: 1385–90).

duty, and (3) that this breach caused some injury to the plaintiff. For simplicity, focus on a case in which there are only two elements, A and B.

If we define the standard of persuasion in terms of individual elements, an obvious difficulty arises: the fact-finder may assess both $P(A)$ and $P(B|A)$ – that is, the probability of B given A – as greater than 0.5 and yet assess $P(A \& B)$, the probability of the conjunction, as being less than 0.5.⁸

An even more serious difficulty is that the number of elements into which a given claim is divided is essentially arbitrary, because the plaintiff's claim may be stated in other ways that are logically equivalent. For example, the proposition that the defendant breached a duty to the plaintiff may be divided into three separate propositions: that some person took conduct of a given general description at a given time and place, that the defendant was that person, and that the conduct was of such a nature as to constitute a breach of the defendant's duty to the plaintiff. Causation and injury are also often thought of as separate elements.

The difficulty is that the more elements a claim is divided into, the easier it is for the plaintiff to satisfy the burden with respect to each element. Suppose, for example, that under one definition of a claim one element of the claim is that no notice of a given fact had been given during a period day 1 to day n , inclusive, and that under a second definition this element is replaced by n elements, each one representing lack of notice for 1 day during the period. Obviously, it is more difficult for the plaintiff to satisfy the 0.5 standard with respect to the first definition than with respect to each of the elements under the second definition. And the difficulty could be diminished further by further subdividing the periods into minutes or seconds. If the time period is divided finely enough, a reasonable fact-finder will necessarily conclude, as to each time period, that it is more likely than not that no notice was given during that period. Thus, defining the standard of persuasion in terms of individual elements becomes incoherent.

The cumulative approach

There is an obvious solution to these problems. In setting the standard of persuasion at 0.5, the basic model that I described above distinguished only between those situations in which judgment for the plaintiff is desirable and those in which it is not, and the former type of situation arises only if both A and B are true. Thus, the proper rule of decision under that model is that the fact-finder should find for the plaintiff only if $P(A \& B) > 0.5$.⁹ Because it is sensible, fact-finders apparently tend naturally to operate in a cumulative manner.¹⁰ I do not believe that any considered body of law precludes implementation of this rule,¹¹ and if there were the Bayesian approach would show that the law was wrong, not the other way around.¹²

8 Often this problem will not arise because the elements of the claim are highly dependent. See, for example, Koehler and Shaviro (1990: 269–70); Lempert (1986: 452). But the problem is always theoretically capable of arising.

9 Where the defendant raises an affirmative defence, a more complex instruction would be required (Lombardero, 1996: n. 55).

10 Cf. Lombardero (1996: 284), summarising research suggesting that juries do not typically analyse the evidence element by element, but rather by testing a variety of hypotheses or 'stories' to explain the evidence.

11 Nance (1986: 947) and Lombardero (1996: 286–7).

12 Hence, although it may be that a defence of the element-by-element approach consistent with Bayesianism might be constructed (see Lempert, 1986: 452–3), I do not believe Bayesians need be defensive about the conjunction matter.

Concerns about the cumulative approach

In a previously published work, Allen has contended, as part of a broad Bayesioskeptical assault, that the cumulative view of the fact-finding process under Bayesian analysis is flawed. As I understand his argument, he raises two basic concerns. Neither, in my view, poses a serious problem.

Weird results One of Allen's concerns is that, '[i]f the dependency relationships among the various elements [of a claim] are not known,' then 'all kinds of weird results obtain' in attempting to determine the probability of combinations of propositions. Most material in this context is that the same piece of evidence may increase the probability of each of two propositions and yet decrease the probability of their conjunction (Allen, 1994: 608). But so what? To produce such results requires a rather unusual (though easily enough imaginable) factual setting. In Figure 1, a Venn diagram, areas represent probabilities; the square in each case itself has area 1. Figures 1a and 1b show, respectively, probability assessments before and after a given piece of evidence is presented. The situation described by Allen as weird arises here because the evidence increases the size of both areas A and B, but pulls them apart to the extent that their overlap decreases.

Suppose that Mom and Dad are both very busy, so that whenever either has errands to do in town he or she takes baby along, leaving the other parent at home to work; baby never comes to town otherwise. Occasionally, though, Mom and Dad leave baby at home with a sitter and come into town for a break together. Then spotting baby in town increases the probability that Mom is in town, increases the probability that Dad is in town, but decreases the probability that Mom and Dad are both in town.

Allen also regards as 'weird' the possibility that a given piece of evidence might increase the probability of two propositions but decrease the probability of their disjunction (Allen, 1994: 608). Again, it requires an unusual setting to create such a result: Figure 2 shows evidence increasing both area A and area B, but pushing them together, increasing their overlap, so much that their total area is decreased.

Suppose Mom and Dad each like to run for exercise, but because the parent not running has to stay at home with baby they each run only every other day. Now they are considering buying a jogging stroller, so that they could take baby along, and if

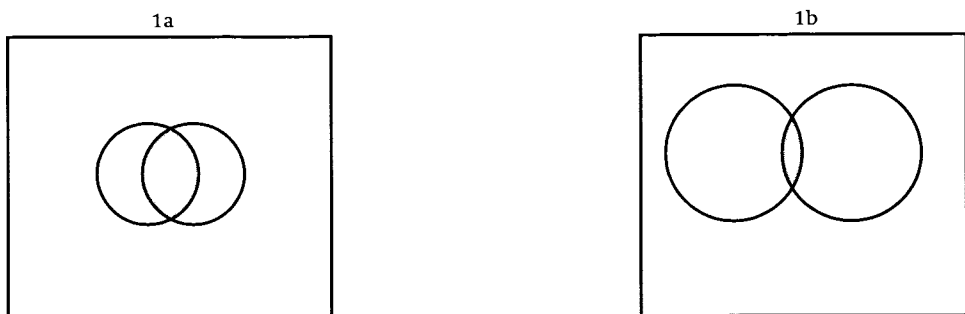


Figure 1. Probability of propositions increases, but probability of conjunction decreases.

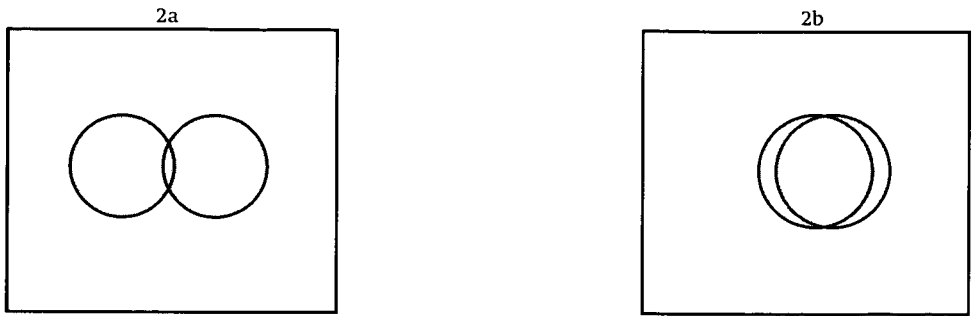


Figure 2. Probability of propositions increases, but probability of disjunction decreases.

they do buy one they plan to take advantage of it by running 5 days per week, nearly always together. Evidence that they have gotten a jogging stroller, therefore, increases both the probability that Mom runs on any given day and the probability that Dad runs on any given day, but decreases the probability that a parent runs on any given day.

Are these results unusual? Perhaps. Weird? I don't think so. But in any event an inferential theory had better be able to handle them; Allen's attempt to use them as the basis for an assault on a probabilistic view of fact finding is unpersuasive.

Now, Allen does condition his argument by suggesting that this problem arises '[i]f the dependency relationships . . . are not known,' but a useful meaning for this qualification is elusive. Of course, the 'dependency relationships' among various propositions cannot generally be known in the sense of absolute certainty, because that sort of knowledge is virtually always unattainable (I am tempted to delete the 'virtually', but maintain it for fear of creating a paradox). On the other hand, if the bearing of evidence *E* on hypothesis *H* may depend on condition *C*, then clearly it will help if the fact-finder has information as to whether *C* is true and on how *C* may affect the *E*-*H* relation. Thus, to assess the probability that Mom is running given proof that Dad is running, it will help to have information as to whether they have a jogging stroller and (if they do) on how having one would affect their routines. This is neither surprising nor particularly troublesome; there may, of course, be some cost to providing the information, but that is true of all information.

Fortuity Allen's other contention is that the cumulative view still makes the standard of persuasion depend on the number of elements with which a claim is defined.¹³ I do not believe it does. Suppose that a given claim ordinarily defined to comprise *m* elements is redefined, without altering the substance, to comprise a greater number *n* of elements. It is true that, given the first definition, the *average* probability for each element must exceed $(0.5)^{1/m}$ for the plaintiff to succeed, while under the other

¹³ Allen (1994: 607) and Allen (1986: 407).

definition the average must exceed a higher number, $(0.5)^{1/n}$.¹⁴ That is of no concern, however. The average probability for each element is of no significance given the cumulative nature of the fact-finding task; a little less probability on one element can be compensated for by a little more on another element.

Moreover, it is clear that, assuming a claim is not altered substantively, dividing the claim into more elements will in fact raise the average probability that a fact-finder would assign to each element. Because the redivision of the claim has not altered its substance, the fact-finder's assessment of the probability of the truth of the entire claim cannot have changed; it follows that the average of the probabilities that the fact-finder assigns to each element must rise. Looked at another way, a corollary of the division of a claim into more elements without altering the substance of the claim is that there is less content in each of the elements. The average probability of the elements would therefore be expected to rise, especially given that the elements are not independent. Taking again the case of an element of no notice broken down into n elements, one for each day in a period, a high probability of lack of notice on day 1 would tend to suggest the same for day 2, only slightly less strongly for day 3, and so forth.¹⁵

Now compare two claims that *are* different substantively, claim B requiring all the elements that claim A does but some additional ones besides. Given the same burden of persuasion, the claimant must prove the elements to a higher average probability to succeed on claim B than to succeed on claim A. Again, though, the average is a figure of no significance. More to the point, the addition of extra elements makes the claimant's burden more difficult on claim B than on claim A. But of course it does, given that claim B requires proof of more substance than does claim A and tolerates no more uncertainty. There is no conceptual difficulty here.

Thus, the so-called problem of conjunction is illusory. There is no logical reason why the fact-finding function ought not to be conceived in cumulative terms. I think there are other types of justification that might be offered for defining the standard of persuasion in terms of individual elements, one based on the difficulty that the adjudicative system has in acknowledging uncertainty,¹⁶ and the other based on

14 Actually, the probabilities used in computing these averages are the probability of the first element, the probability of a second element given the first element, the probability of a third element given the first two elements, and so forth.

15 I am taking a somewhat different approach from that in Dawid (1987: 95–6). I do not believe Dawid's conditions – the likelihood ratio insensitive to the substance of the elements, and the number of witnesses varying with the number of elements – are helpful, and they gratuitously left him open to challenge by Cohen (1988).

16 Because of the gravity of its decisions, the adjudicative system seems to prefer acting in the end as if truth has been determined with certainty. That is possible under an element-by-element approach to the standard of persuasion. If all elements have been determined in favour of the plaintiff, then it is taken that his claim is valid and he wins; otherwise, he loses. Under a cumulative approach, however, this is not possible. See Nesson (1985: 1390). The fact-finder may be persuaded that each individual aspect of the plaintiff's claim is true, and yet conclude that the aggregate uncertainty is sufficiently great that the plaintiff ought to lose. If the fact-finder is articulate, the cumulative definition may prevent it from being able to pinpoint a reason for denying relief to the plaintiff.

efficiency.¹⁷ I do not believe that these arguments are persuasive.¹⁸ But even if they are, they should be recognised as policy-based grounds to depart from a logical approach to fact finding at the level of the overall case.¹⁹ They do not suggest that an approach to fact finding based on the conventional theory of probability is illogical.

17 Assuming, as we have, that given elements of a claim are conjunctive, so that the plaintiff must prevail on all of them to win the case, a convenient short-circuit may be available: if the fact-finder determines that any one of the elements is more likely to be false than true, a verdict for the defendant is appropriate without further ado. So far, that is not a problem for the cumulative view of the fact-finding process because a suitable instruction could express this shortcut. Now suppose, however, that the court is considering the possibility of making the trial sequential, so that one element is tried before the others. If, for example, it appears very likely that the fact-finder will find against the plaintiff on element A, that element B will be long and costly to try, and that there is not too great an overlap in the evidence to be presented on the two elements, it is a tempting solution to try element A first: a finding that element A is more likely to be false than true would obviate the need to try element B, whether the fact-finding function is viewed element-by-element or cumulatively. But if the fact-finder finds that element A is more likely to be true than not, then element B must be tried. Under the element-by-element approach, this poses no problem: element A is taken as proven, and now the question is whether element B is more likely than not. Under the cumulative approach, though, we do have a problem. The fact-finder's task at this point is not simply to determine whether B is more likely to be true than not; it must still determine whether A & B is more likely true than not. That determination continues to depend in part on the probability of A, and by the time the evidence on B has all been presented to the fact-finder the evidence on A, and even the degree of the fact-finder's own degree of confidence in A, may be a rather distant memory. In some cases, the delay may be so great that there is a new fact-finder altogether.

18 Whatever weight reluctance to acknowledge uncertainty may have, it is not enough to warrant rejecting a sensible view of the fact-finding process for one that is neither sensible nor coherent. Courts should simply accustom themselves to the fact that they cannot act as if a factual finding of a proposition renders that proposition certain, and that sometimes cumulative uncertainty may undermine a claim even though each individual element of the claim appears more likely to be true than not. Put another way, although if A and B are each true it follows logically that A & B is true, if A and B are each probable it does not necessarily follow that A & B is probable, and the legal system should not act as if it does. Moreover, courts should not be mesmerised by an attempt to determine 'the single most probable story'. As suggested below, there is no clear boundary line to a story; to make the fact-finding process coherent, a wealth of infinitesimally different story lines must be batched together. As for the possibility that sequential trials are simpler under an element-by-element view of fact finding than under a cumulative view, it seems that this is true only if one ignores the incoherence introduced by the element-by-element view, whether the trial is sequential or not. The problem of holding a sequential trial under the cumulative view might perhaps be ameliorated if, assuming that at the conclusion of the first stage the fact-finder believes A to be more likely than not, it attempts to articulate its degree of confidence in A. If the fact-finder were able to assign a numerical probability P to this probability, then the plaintiff would not prevail unless at the second stage she persuaded the fact-finder that the probability of B given A was greater than $1/2P$. Assuming the fact-finder cannot operate in numerical terms in this way, it might still state its degree of confidence verbally, using terms like 'only slightly more likely than not', 'very confident', or 'virtually certain'. In any event, the problem of sequential trials is a relatively minor one, because in the overwhelming majority of cases courts do not even consider splitting the trial.

19 There is an inverse question of whether, or when, courts ought to be willing to cumulate related but different claims in favour of a plaintiff, for example, saying in effect, 'Claim A is less than 0.5 probable, and so is claim B, but it's more than 0.5 probable that A or B is accurate, so we'll give you judgment'. Different considerations may apply to that question, and I put it aside for purposes of this paper. See *Schad v Arizona* 501 US 624 (1991); Hodgson (1995: 62-3); Lombardero (1996: 288-9).

The presumption of innocence

Along with the standard of persuasion, another part of the burden of proof is what is commonly called the burden of production, the designation of which side (if either) loses with respect to a given proposition if nobody presents any evidence at all or an insufficient degree of evidence. Another way of thinking about this burden is as an initial presumption. In criminal cases, especially, we often speak of the presumption of innocence, meaning that the prosecutor or other claimant has the burden of production. The presumption is policed in part, when there is a jury, by keeping the case away from the jury if the claimant has not satisfied the burden of production. But if the case does go to the jurors, they must still be told about the presumption so that they know what their point of departure should be in considering the evidence. Bayesioskeptics have contended that the presumption of innocence is inconsistent with a nonzero prior probability of guilt (see *State v Skipper* 228 Conn. 610, 637 A. 2d 1101 (1994); Cohen, 1977: 109–13; Bergman and Moore, 1991: 597–8; Tribe, 1971: 1361). A zero prior probability of guilt represents a belief in innocence that cannot be shaken no matter how strong the evidence.

I believe this argument is wrong, but I do not believe the easy answer that has been offered by Bayesians is quite satisfactory. The easy answer in a criminal case is to say that the presumption means that, before any evidence is presented, the fact-finder should treat the defendant as no more likely than anyone else, or, in a variant, as equally likely as anyone else, to be guilty.²⁰ I have at least four problems with this test. First, it is indeterminate: there is no obvious single reference population.²¹ Second, the defendant may contend that he is substantially *less* likely than the 'mean' member of the population to have committed the crime. Third, information that the fact-finder may legitimately bring into court, without the need for evidence, may suggest that the defendant is *more* likely than most other members of the population to have committed the crime; consider, for example, the outset of a case in which the jury knows nothing but that the charge is armed robbery and that the name of the defendant is John (suggesting that he is male and therefore more likely to have committed armed robbery than the average member of the population). Finally, and perhaps most significantly, this test only touches on the identity element of the charge or claim. The defendant may not concede that any crime or other wrong was committed at all, and unless he does so the claimant (the plaintiff or prosecutor) has the burden of production on that issue as well as on the question of identity.

So I think that conceptualising the presumption of innocence in Bayesian terms is trickier than has been acknowledged by good Bayesians. I think it can be done, though. The presumption means in effect that, before considering the evidence, the fact-finder should use a very low probability for the proposition that the defendant is

20 See the comment of Kaye and Balding in Allen *et al.* (1995: 293), expressing the no-more-likely version, and n. 13 quoting Dawid, expressing the equally-likely version.

21 See Robertson and Vignaux (1991: 641 and n. 36), saying that '[i]t is not clear that there can be any general rule as to the prior probability of guilt in a criminal case', and that to use as a reference population a group such as the population of the greater Los Angeles area 'represents an artificial truncation of the ordinary reasoning process'.

guilty, the probability that the defendant would implicitly use if the question had never been raised.²²

Individual items of evidence

So far I have examined the overall task of the fact-finder with respect to the entire body of evidence. But conventional probability theory also offers insights on assessing the weight of individual items of evidence. At the same time, this function has also led to various objections by the Bayesioskeptics.

Story-telling and Bayesianism: a false dichotomy

Recent research suggests that fact-finders tend to view an entire body of evidence, attempting to determine a story that most plausibly accounts for all of it (e.g. Pennington and Hastie, 1991: 529). Some Bayesioskeptics have viewed this model as conflicting with a Bayesian view of evidence.²³ I do not believe there is any genuine conflict (Friedman, 1992: 96–7).

The perception of a conflict, I believe, is attributable to a misconception about the role of Bayes' theorem in Bayesian reasoning. The theorem is important, but it is not all of Bayesianism.²⁴ Let us think about the theorem at its simplest. Suppose there are two mutually exclusive and collectively exhaustive hypotheses, H_1 and H_2 . Information already received, including background information about the world, leads the fact-finder to assess $P(H_1)$, and with it $P(H_2)$, which necessarily equals

22 This brief statement may cover up some ambiguity. If the question was never raised the fact-finder may not have had to make any decision that depended on an assessment, even implicit, of the probability of that proposition. On the other hand, if the question were raised, the very posing of it – especially, but not only, in a formal charge – tends to carry the suggestion that the proposition it describes is at least plausible. I think, though, that a thought experiment suggests the type of assessment I mean. Imagine that the fact-finder, before being presented with any evidence bearing on this case, is presented with a very, very long questionnaire, containing all sorts of propositions, many of them extremely unlikely because they describe very particularised outcomes:

How likely is it that there were 3.49999 inches of rain as measured in City Park last month? 3.49998? 3.49997? ... That Victor Varmint was murdered by the first person listed in the Muddletown 'phone book? The second person? The third? ... That the next winner of the state lottery will be the first person listed in the Muddletown 'phone book? The second person? The third? ...

I am not suggesting that a juror ought to be presented with such a questionnaire, or even with this thought experiment, but only that it suggests an intellectually coherent way of thinking about the presumption of innocence in Bayesian terms. For earlier thoughts about the same problem, including a jury instruction that might suffice, see my comments in Allen *et al.* (1995: 296).

23 For example, Ligertwood has said that Bayesians 'would have us conceptualising our beliefs in mathematical terms and then increasing those beliefs by adding in, piece by piece, individual items of evidence, also conceptualised as mathematical chances, until our beliefs have achieved a degree of likelihood we are prepared to accept as proof'. (Unpublished paper quoted in Robertson and Vignaux, 1993: 475).

24 Robertson and Vignaux (1991: 633); Robertson and Vignaux (1993: 476). I have suggested near the beginning of this article that application of the conventional probability theory to subjective probability assignments is the essence of Bayesianism; Bayes' theorem is an important aspect, but not all, of the standard probability theory.

$1 - P(H_1)$. Now a new piece of evidence E is received.²⁵ E is trace evidence; it arose after the event that is the subject of H_1 and H_2 . Thus, the fact-finder can assess $P(E|H_1)$, that is, the probability of E given H_1 (as well as all the other information received) and $P(E|H_2)$. But this is not what she needs to know; instead she needs to know $P(H_1|E)$. Bayes' theorem shows how to transpose the conditional:

$$P(H_1|E) = \frac{[P(H_1) \cdot P(E|H_1)]}{[P(H_1) \cdot P(E|H_1)] + [P(H_2) \cdot P(E|H_2)]}$$

If another piece of evidence is now presented, Bayes' theorem can be applied again, with $P(H_1|E)$ now acting as the prior probability. And so on indefinitely – a meal of evidence ingested and digested one bite at a time. But note how complex the problem has become already; the prior probability, instead of being a simple term, is an ungainly fraction, and it must be used to generate a far more awkward fraction – and so on indefinitely.

I do not believe that fact-finders can, should or do go through such a serial updating of probability, given each new piece of evidence. Thus, I believe the court in the recent case of *R v Adams* [1996] 2 Cr App R 467 – though other aspects of the opinion strike me as wrongheaded – might have had a point in saying that the attempt to lead the jury through this kind of process was inappropriate.

Moreover, not all pieces of evidence call for application of Bayes' theorem. I have spoken of trace evidence, which arises after the events in dispute. But now consider evidence that arises *before*, such as evidence of motive. The key is not chronology so much as causation. The evidence may, for example, be proof of a motive that might have led to the crime. The fact-finder is likely to begin with some assessment of how likely the crime was given the motive. There is no need to assess how likely the motive (evidence) was given the crime (hypothesis), just so the transposition of Bayes' theorem may be performed.

How, then, can a Bayesian approach operate on a whole body of evidence, some of which may arise before the events at issue and some after? Suppose the fact-finder, in choosing between H_1 and H_2 , has four pieces of evidence, E_1 having arisen beforehand and E_2 , E_3 , and E_4 having arisen afterwards. Then the fact-finder may compare competing stories, one for H_1 and one for H_2 , that account for all of the evidence. That is, the fact-finder may compare the relative probability of:

$$P(H_1|E_1, E_2, E_3, E_4) = P(E_1) \cdot P(H_1|E_1) \cdot P(E_2|E_1, H_1) \cdot P(E_3|E_1, H_1, E_2) \cdot P(E_4|E_1, H_1, E_2, E_3) \quad (4)$$

and

$$P(H_2|E_1, E_2, E_3, E_4) = P(E_1) \cdot P(H_2|E_1) \cdot P(E_2|E_1, H_2) \cdot P(E_3|E_1, H_2, E_2) \cdot P(E_4|E_1, H_2, E_2, E_3) \quad (5)$$

The comparison is a little simpler yet, because $P(E_1)$ is common to each of these equations, and so drops out of the comparison.

This is a story-telling model, but it is also Bayesian. A careful analyst using this model to compare competing stories will focus on some of the individual linkages, for

²⁵ I do not mean to suggest in this discussion that there are any prescribed bounds to a 'piece' or 'item' of evidence. It may, for example, be most convenient to think of two separate witnesses testifying to the same fact as one piece of evidence.

example, 'How probable is it that E_3 would arise given E_1 , H_1 and E_2 , and how probable given E_1 , H_2 and E_2 ?'

In contending that Bayesian reasoning does not always depend on applying Bayes' theorem to individual items of evidence, I do not mean to suggest that such an application lacks value. On the contrary, the theorem is very useful in establishing a framework for analysing the probative value of a piece of evidence, or of a category of evidence. That is not the job of the fact-finder; the fact-finder's job is to determine the bottom line, the probability of the facts in dispute, and not the value of individual items of evidence. But it is precisely this latter determination that courts must perform as part of the task of determining the admissibility of a piece of evidence, and that rule makers must perform in crafting admissibility rules. Bayes' theorem does not purport to force item-by-item analysis on the legal system; it is the legal system that sometimes demands such analysis, and the Bayesian approach is able to respond.

Computational complexity

Some Bayesioskeptics emphasise the computational complexity created by the Bayesian approach.²⁶ They argue that it is beyond the powers of the human intellect to take into account and reconcile all the probabilistic assessments that full application of the Bayesian approach requires when applied to a situation of even moderate complexity.

I think the argument is wide of the mark, for several reasons. First, the world is indeed a complex place, but that does not reflect a problem with Bayesian analysis (Friedman, 1992: 95). On the contrary, any theory that could not in principle represent the complexity surrounding us would have limited value.

Second, I say 'in principle' because the theory need not be applied in its most powerful gear. On the contrary, it is a flexible template. It can take into account as much complexity as its user is able to handle. If constraints on the user's capacity mean that – at the cost of ignoring information and so altering the probabilistic assessments to some extent – she must distill much of the complexity out of the situation, the theory may still be applied. I believe that we do simplify factual situations to make them tractable. One way we do so is by batching various possibilities together.

For example, suppose that in a murder case the evidence tends to prove that the time of death was around 10:25 p.m. and that the defendant was seen at 10:40 p.m. at a place that one would expect to take 20 minutes to reach. Even with this simple evidence, one can identify an enormous, and perhaps infinite, number of possible stories consistent with the defendant's guilt. Each nanosecond has its own story: The earlier before 10:25 p.m. that the murder occurred, the less likely the time-of-death evidence is to have arisen, but the more likely the defendant would have been able to reach the distant place by 10:40 p.m. But of course we do not analyse each nanosecond's story on its own; rather we batch large groups of them together.

Similarly, suppose that the identity of the murderer is in issue. Then the defence has an enormous number of alternative hypotheses – each human being who may then have been alive provides one (actually, a multiplicity of them, but they may be batched). Of course we need not consider each individual on his or her own; we batch

²⁶ This is a point that has been especially emphasised by Callen (see Callen, 1982: 10–15; Callen, 1986: 726, n. 72; Callen, 1992: 1405; Callen, 1991: 1118, n. 19; Callen, 1994: 76, n. 50).

them together in large groups,²⁷ and some of the alternative hypotheses – people who live at the other end of the globe – may be so slight that they can be disregarded altogether.

Third, I think the computational complexity, though great, may not be as great as it appears. The storytelling Bayesian approach that I have outlined above is far simpler than an item-by-item approach. It requires a limited number of probability assessments, and does not demand that the observer cross-check for consistency all the probabilities that she might assess given multiple items of evidence. Also, I do not think pure computational precision or consistency is necessary to make the Bayesian system work. Here, tentatively, are the sketches of two models that I suspect might be useful, individually or integrated together.

One model requires simply that an observer adhere to the following rule: If $P(E|H) > P(E|NH)$, then $O(H|E) > O(H)$, and the greater the proportional difference between the first pair the greater the proportional difference between the second pair. This rule follows immediately from Bayes' theorem. One could comply with this rule without complying with the theorem, but usually the rule would tend to lead to results rather close to results prescribed by the theorem. The rule expresses the commonsense proposition that if a piece of evidence is more likely to arise given an hypothesis than given its negation, the evidence makes the hypothesis more likely.

In the second model, an observer might at any time make a set of probability assessments that are inconsistent. If the inconsistency is a glaring one, the cognitive dissonance will be apparent, and the observer will adjust the assessments to bring them more closely into line. This adjustment might involve altering any probability assessment, including a prior probability.²⁸ The tendency will be to move towards equilibrium, in which there are no inconsistencies, but new information might disrupt the situation and start the process again.

Finally, the computational complexity problem really only arises to the extent that Bayesians argue that fact-finders do or should perform their task by working through a Bayesian calculation.²⁹ At least for the most part, this is not the claim. Rather, I believe two claims are justified.

First, when thinking well – and with the aid of whatever simplifications and heuristics may be necessary – fact-finders reach results that are roughly consistent with those they would reach if they were to apply probability theory rigorously. They do not have to think about the theory consciously, just as an athlete does not have to think about the laws of physics in determining where a ball hurtling through the air is likely to land on the ground. But just as athletes manage to predict the path of the ball quite well, most of us manage to handle uncertainty in our daily lives to a tolerable degree; we do not stay home paralysed by fear of the unknown, and we do not get run over by the fast-moving vehicles that come near us.

Second, probabilistic analysis can help observers – lawyers, courts, and especially scholars – think about the probative value of evidence. As David Kaye has put it, the

27 See the response by Kaye and Balding to Allen's 'phone book' argument in Allen *et al.* (1995: 294).

28 That is, the observer might in effect think something like: 'I thought $P(H_1)$ was very low, but now that I have received evidence E I think that $P(H_1|E)$ is higher than would be justified by my prior assessment of $P(H_1)$ and my assessments of $P(E|H_1)$ and $P(E|H_2)$. So something is out of kilter, and in part I think it's that I assessed $P(H_1)$ too low'. For earlier discussions of this idea, see my comments and those of Kaye in Allen *et al.* (1995: 299, 306).

29 See Hodgson (1995: 54–7, 62) arguing that such an approach is 'impossibly complex'.

proof of the value of using probability theory as an analytical tool lies in the pudding, and I am happy to agree with him that this approach 'has produced some treats' (comment in Allen *et al.*, 1995: 305).³⁰ I will offer one rather simple illustration: Bayesian analysis shows that there is more value than has recently been thought to the rule precluding experts from testifying to the 'ultimate issue,'³¹ and suggests that often the preferred form of expert testimony is in terms (whether explicit or not) of a likelihood ratio: 'This evidence is x times more likely to have arisen under Hypothesis 1 than under Hypothesis 2' (Robertson and Vignaux, 1995: 60–5).

Indeed, I cannot recall the Bayesioskeptics ever offering any criticism about *particular* uses of probabilistic methods as a tool for analysing evidentiary questions; the challenges always seem to be at the general level, concerning the value of the enterprise itself or the overall standard of persuasion. I think they should examine the particulars of what we are doing. I claim that probabilistic methods have helped us achieve results that are not obvious but that are sound, intuitively appealing, and readily explainable. If I am right, the Bayesioskeptics should acknowledge that. If I am wrong, they should show why. Meanwhile, we will continue happily cooking away.

Explicit presentation of probabilistic methods in court

I have argued that the value of Bayesian analysis of evidence does *not* depend on the assumption that probability theory will be explicitly presented in court, through expert testimony, argument of counsel, or (where there is a jury) instruction by the court. But because the question of explicit presentation arises recurrently – very recently and dramatically in the *Adams* case – I will briefly address this question.

Unless statistical evidence plays a significant role in a case, I believe there is usually no substantial reason to make an explicit presentation of probability theory; fact-finders can deal with the evidence much as they deal with ordinary questions in their everyday lives. On the other hand, I believe there is no longer any real general objection to the presentation of statistical evidence,³² though in some settings even devout Bayesians (including me) have found appeal in verbal, rather than quantified, presentations (Robertson and Vignaux, 1995: 56–7, presenting a set of verbal numerical equivalents proposed in Evett, 1991, and Aitken, 1995: 52, presenting a somewhat different, and perhaps more satisfactory, set proposed in Evett, 1987: 103).

Assuming that statistical evidence is to be presented, the question is what explicit explanation of probability theory ought to be offered. I think this is a practical, rather than a theoretical, issue. Assume that the components of the likelihood ratio of the statistical evidence, or the ratio itself, will be presented. If we think jurors are able to make reasonably good use of this evidence without a tutorial, perhaps we should do without one, content to present the evidence and let them do the rest. If, furthermore, we think that, however incompetent the jurors may be, a tutorial is likely to confuse them more than enlighten them, then we should probably do without it. But in some circumstances we might decide that some tutorial would be of net assistance.

30 This is also a clearly self-interested view on my part, given Kaye's kind comment on the contribution of what he generously – and too modestly – calls the 'Michigan School'.

31 The rule has been largely abrogated for the federal courts in the United States by Fed R Evid 704, and essentially discarded in England by *R v Stockwell* [1993] 97 Cr App R 260.

32 Wright (1988) may disagree. The Minnesota Supreme Court has at times seemed to adopt something close to this view, but more recently it has, sensibly, begun to back off. See *State v Bloom* 516 N.W. 2d 159 (Minn. 1994).

In the case of DNA evidence, for example, it *might* be helpful to show with the use of Bayes' theorem just how powerful the evidence is, that is, what effect it should have in altering a juror's assessment of the probability of guilt. True, this approach focuses on one item of evidence, but the fact is that DNA evidence often deserves the focus, for at least two reasons. First, it can be extremely powerful. Second, because it is statistical in nature and tends to involve very large numbers (or very small ones, depending on how they are expressed), it can be especially difficult to use. Thus, I think the *Adams* court was off base in suggesting that Bayes' theorem has no place in the courtroom.

On the other hand, as I have already indicated, I tend to think that it is not helpful to take a jury through an iterative use of Bayes' theorem, even in a case like *Adams* in which the evidence was very sparse. I acknowledge that the defence faced a significant problem: it had to show the jurors, in effect, how low they might rationally assess the prior probability to be, so that they did not treat the DNA evidence, on which the prosecution depended virtually exclusively, as tantamount to a certain proof of guilt. It seems to me, though, that it would have been better, if counsel believed expert help was necessary on this point, to focus directly on it, rather than using the expert to suggest to the jurors that they employ an iterative algorithm that almost certainly appeared to them to be unrealistically mechanical as well as utterly bewildering.³³

I do not claim any great certitude on this point; the problem of presentation is a tenacious one. We can hope that before long technology will obviate the problem so far as DNA evidence is concerned, and that this evidence can be presented as simply as other forms of scientifically based evidence, such as the readings of radar guns, for which we never think statistics are necessary.

Conclusion

It is necessary to keep Bayesian methods in their proper place with respect to juridical proof. For the most part, they are of analytical assistance only, to those who think about and craft evidentiary law – but for that purpose they are of very great assistance indeed. For the most part, probability theory should not be presented to the fact-finder, even though it helps shape the evidence that the fact-finder receives. In some settings, perhaps, aspects of the theory should be presented, but I do not have great hopes that it will do much good: I believe that the courtroom is not the place to compensate for failures in prior learning.

If the limited role of Bayesian analysis is kept in mind, then I believe that the arguments of the Bayesioskeptics lose most of their force. Probability theory is a useful, even inevitable, way of thinking about evidentiary law. But, just as it is generally best that the audience not see what is going on backstage, I do not believe that probability theory usually needs to be mentioned in the courtroom.

33 Counsel might ask, 'Please assume these items of evidence: A, B, C, . . . N. Now, is there a set of assumptions on which a reasonable person, considering only this evidence and not the DNA evidence, might assess the probability that the accused was at the scene of the crime as less than $1/X$? Can you state one or more such sets?'